Reproducibility and Validity of Radioimmunoassays for Urinary Hormones and Metabolites in Pre- and Postmenopausal Women

Roni T. Falk,[1] Mitchell H. Gail, Thomas R. Fears, Susan C. Rossi, Frank Stanczyk, Herman Adlercreutz, Paula Kiura, Kristiina Wahala, Jennifer L. Donaldson, Jimmie B. Vaught, Capri-Mara Fillmore, Robert N. Hoover, and Regina G. Ziegler

Division of Cancer Epidemiology and Genetics [R. T. F., M. H. G., T. R. F., J. L. D., C-M. F., R. N. H., R. G. Z.] and Division of Cancer Prevention and Control [S. C. R.], National Cancer Institute, Bethesda, Maryland 20892; Infertility and Endocrinology Department, Women's and Children's Hospital, University of Southern California, Los Angeles, California 90033 [F. S.]; Department of Chemistry, University of Helsinki, SF-00290 Helsinki, Finland [H. A., P. K., K. W.]; and Microbiological Associates, Rockville, Maryland 20850 [J. B. V.]

Abstract

The reproducibility of RIAs of circulating sex hormones has been evaluated as part of recent epidemiological investigations, but none seem to have addressed the reproducibility or validity of RIAs for urinary hormones or their metabolites. As part of a case-control study of breast cancer in Asian-American women, 12-h overnight urine samples were obtained, and a methodological study was conducted to identify laboratories capable of assaying urinary hormones. For the reproducibility component of this study, two laboratories with extensive experience in hormone assays measured urinary estrone, estradiol, estriol, pregnanediol glucuronide, and estrone glucuronide using samples from 15 women (5 midfollicular, 5 midluteal, and 5 postmenopausal). Variance estimates from these measurements were used to calculate the laboratory variability (coefficient of variation) and to assess the magnitude of the biological variability among the women in relation to the total variability (intraclass correlation coefficient). For the validity component, urinary estrone, estradiol, and estriol levels were measured in the same samples by gas chromatography-mass spectroscopy in the laboratory of Dr. Herman Adlercreutz (University of Helsinki, Helsinki, Finland). We found that the degree of assay reproducibility differed between the laboratories, but that laboratory variability was usually low compared with the range of hormone values among women, particularly for the estrogens. Values for estrone and estradiol were well correlated among all of the laboratories. For estriol, the RIAs tended to overestimate levels compared with gas

chromatography-mass spectroscopy. In one laboratory, assays for pregnanediol glucuronide and estrone glucuronide were consistently reproduced; in the other, the reproducibility of the RIA for pregnanediol glucuronide was problematic, and estrone glucuronide was not measured. Despite some limitations, urinary hormones and their metabolites can be reliably measured by current RIAs in large investigations attempting to link hormone level to disease risk and may be particularly advantageous for studies of postmenopausal women, where serum concentrations of estrone and estradiol are low and assay measurements are not as dependable.

Introduction

Urine samples have long been used to characterize clinically relevant levels of steroid hormones, particularly for monitoring pregnancy progression (1, 2). In epidemiological studies, the collection of urine has several advantages compared with blood samples: collection is noninvasive and can be conducted by the subject without the assistance of a health professional, and samples can be easily stored and transported. Measurements reflect hormone levels over a period of time, typically 12 or 24 h, which may cover several cycles of pulsatile secretion. Moreover, because the concentration of most urinary hormone metabolites is high, costly extraction procedures needed for blood assays are not required. Concerns about reproducibility within and comparability among laboratories using RIAs to measure serum hormones have been raised in large population-based studies (3–7), but little attention has been paid to the quality of urinary hormone measurements. Problems with assay reproducibility compromise the power to link hormone levels with cancer risk; when measurements are reproducible, but do not reflect the true hormone values (i.e., the estimates are biased), study results may be difficult to interpret.

Plasma and 12-h overnight urine samples were obtained as part of a case-control study of breast cancer in Asian-American women to study the role of endogenous hormones. Before analyzing these specimens, a methodological study was conducted to identify laboratories that could reproducibly and accurately measure circulating and urinary hormones and their metabolites. Findings on the reproducibility of plasma estrogens and progesterone have been presented elsewhere (7). In this study, we present results of urinary hormone measurements, using 12-h overnight urines collected from 15 volunteer women. Each woman's urine was assayed repeatedly over the course of 4 months. To estimate "true" hormone values, urinary estrogens were measured by GC-MS[2] methods.

[2] The abbreviations used are: GC-MS, gas chromatography-mass spectroscopy; CV, coefficient(s) of variation; ICC, intraclass correlation coefficient.

Materials and Methods

The design and analytic methods for the reproducibility study have been presented in detail elsewhere (7, 8). To summarize, we obtained 12-h urines from a total of 15 female volunteers not currently using exogenous hormones: 5 were in the mid-follicular phase of their menstrual cycle (6–10 days after the start of menses; mean age, 40), another 5 were in the midluteal phase (4–6 days before the start of the next menses; mean age, 39), and 5 were postmenopausal with at least 3 years since their last menstrual cycle (natural menopause, mean age, 56). To confirm menstrual phase, premenopausal women were contacted regarding the date of their subsequent menses.

Urine was collected overnight (12 h) into a half-gallon container with a teaspoon of boric acid as a preservative. Containers were kept in ice or refrigerated until the following day when urines were decanted and aliquotted into 15-ml conical tubes and stored at −70°C. For the reproducibility component of the study, two laboratories with considerable experience in hormone assays (one academic, one commercial) received four batches of urines at one time, each containing two blinded tubes per subject, and were instructed to assay one batch per month at the start of 4 consecutive months. Samples were stored at -70°C until assayed. Tubes in each batch were placed randomly and in a different order every month. Although unaware which samples corresponded to which woman, the two laboratories were told whether the sample came from a pre- or postmenopausal woman to facilitate sample handling. Tests were done in duplicate; thus, for each woman, a total of 16 measurements of each urinary hormone was obtained. The laboratories were instructed to assay only those analytes for which they had experience and standardized assays. Both laboratories measured estrone, estradiol, estriol, and pregnanediol glucuronide; only one measured estrone glucuronide. Results for 2-hydroxyestrone and 16-α hydroxyestrone have been presented elsewhere (8).

For the validity component of the study, a single aliquot of urine from each of the 15 women participating in the study was sent in one batch to the laboratory of Dr. Herman Adlercreutz (University of Helsinki, Helsinki, Finland). The laboratory was told whether the sample came from a pre- or postmenopausal woman. Samples were stored at -70°C until analyzed. For each woman, GC-MS analysis was performed in duplicate, and measurements of urinary estrone, estradiol, and estriol were provided.

Laboratory Methods

Laboratory 1 (RIA). For the estrogens, urines were hydrolyzed with β-glucuronidase, followed by solid phase extraction and celite chromatography fractionation to separate the analytes, which were then measured by RIA. Reported intra-assay CV were 5.9%, 6.7%, and 10% for estrone, estradiol, and estriol, respectively; corresponding interassay CV were 6.9%, 8.9%, and 13.5%. All assays were reported to be highly compound-specific, with cross-reactivity with other steroids being <0.001%. Lower limits of detection for the analytes were as follows: estrone, 10 μg/liter; estradiol, 0.2 μg/liter; and estriol, 2.5 μg/liter. The average sample recovery was 90%, and samples with <50% or >105% recovery were repeated. For pregnanediol glucuronide, study specimens and quality control pool specimens were diluted 1:500 with assay buffer, followed by RIA (9, 10) which was incubated for 16 h at 4°C. The sensitivity of the assay was 0.01 mg/liter. Samples with levels >3 mg/liter were repeated with further dilution. The laboratory reported intra- and interassay CV of 6% and 9.8%, respectively.

Cross-reactivity with pregnanediol was 6.7% and <0.01% with other steroids. The assay was acceptable if at least two of the three QC pool values were within two SDs of the mean.

All RIAs were done in duplicate. If the CV was >20%, the sample was repeated.

Laboratory 2 (RIA). Urinary estrone, estradiol, and estriol were quantified by RIA. After their hydrolysis in urine (1 ml) with β-glucuronidase/arylsulfatase and selective extraction, first using diethyl either (extracts estradiol and estrone) and then 40% ethyl acetate in hexane (extracts estriol). The extract containing estrone and estradiol was then subjected to Celite column partition chromatography (stationary phase: ethylene glycol). Estrone was eluted in 15% ethyl acetate in isooctane, and estradiol was eluted in 40% ethyl acetate in isooctane. Similarly, estriol was purified on a Celite column (stationary phase: methanol/water (60/40). Elution of estriol was achieved with 30% ethyl acetate in isooctane. After their purification, the three estrogens were quantified by specific RIAs, as described previously (11–13). Efficiency of hydrolysis was monitored by using external standards of [³H]estrone sulfate and [³H]estrone glucuronide. Procedural losses were monitored by use of [³H]estrone, [³H]estradiol, and [³H]estriol as internal standards, which were added after the hydrolysis step. Creatinine was measured in all urine samples, and the concentration of each estrogen is reported in μg/g of creatinine. The assay sensitivities for the estrone, estradiol, and estriol assays were 0.4 ng/ml, 0.02 ng/ml, and 2 ng/ml, respectively, and the intra-assay and interassay CV ranged from 5–10% and 10–15%, respectively, for each of the three assays. Estrone glucuronide was quantified in urine by direct RIA, after the urine was diluted 1:500 with assay buffer, as described previously (9). The assay sensitivity was 0.02 μg/ml, and the intra-assay and interassay CV were 5.2% and 12.1%, respectively. Measurement of pregnanediol glucuronide in urine was also achieved by direct RIA after dilution of the urine with assay buffer (1:5000), as described previously (9).

GC-MS Method. After the hydrolysis of conjugates, isotope dilution GS-MS in the selected ion monitoring mode was used to identify urinary estrogens (14). A total of 14 estrogens, including E₁, E₂, E₃, 2-hydroxyestrone, 2-hydroxyestradiol, 2-methoxyestrone, 2-methoxyestradiol, 4-hydroxyestrone, 15α-hydroxyestrone, 16α-hydroxyestrone, 16α-hydroxyestrone, 16-ketoestradiol 16-epiestriol, and 17-epiestriol, was measured. Estrogen conjugates were extracted on Sep-PakC₁₈ cartridges and purified on the acetate form of DEAE-Sephadex. The samples were then hydrolyzed using Helix pomatia juice and purified on the acetate form of quaternary aminoethyl-Sephadex. Recovery after hydrolysis was estimated to be 75–82%, based on the addition of deuterated (d₅-)-ethoxime derivatives of all ketonic estrogens as internal standards (15); these deuterated estrogens were later used to correct for these losses. Estrogens with vicinal cis-hydroxyls and diphenolic compounds were fractionated on the borate and bicarbonate forms of quaternary aminoethyl-Sephadex, respectively. Neutral steroids were removed by the freebase form of DEAE-Sephadex, after which estrogens were separated into two groups using Lipidx 5000 in a straight phase system. After trimethylsilyl ether derivatization, estrogens were analyzed by capillary gas chromatography with stable isotope dilution mass spectrometry. Deuterated internal standards were available for all of the estrogens, except 16 β-hydroxyestrone and 17-epiestriol, and were used to correct for losses after the hydrolysis step. However, before the introduction of the deuterated internal standards, it is estimated that 5–10% of the hormones may be lost

Table 1 Geometric mean of urinary hormones by laboratory and menstrual groups: Hormone Feasibility Study, 1995

| | RIA | | GC-MS laboratory | Two-sided signed rank P (overall means)[a] |
	Laboratory 1	Laboratory 2		
Estrone				
Follicular	16.90	17.99	15.52	1 vs. GC-MS, 0.001
Luteal	10.62	10.72	8.65	2 vs. GC-MS, 0.020
Postmenopausal	1.80	1.26	1.38	
Estradiol				
Follicular	7.11	6.44	5.13	1 vs. GC-MS, 0.001
Luteal	4.52	3.66	2.86	2 vs. GC-MS, 0.041
Postmenopausal	0.52	0.39	0.39	
Estriol				
Follicular	15.67	21.53	5.50	1 vs. GC-MS, 0.001
Luteal	16.44	22.34	4.33	2 vs. GC-MS, 0.001
Postmenopausal	3.22	3.87	0.83	
Pregnanediol glucuronide				
Follicular	0.82	1.75		
Luteal	3.43	6.19		
Postmenopausal	0.13	0.28		
Estrone glucuronide				
Follicular		37.67		
Luteal		26.18		
Postmenopausal		2.92		

[a] Laboratory *versus* laboratory.

due to incomplete hydrolysis, and this loss cannot be quantified. The laboratory was shipped 15 aliquots of urines and analyzed each sample in duplicate. These duplicate runs were not blinded. For this analysis, the duplicate results were averaged. For all of the estrogens, the limit of detection varied from 0.5–3 nmol/liter. The CV in premenopausal urine samples for the 10 major estrogens were reported to range between 4–7%.

Statistical Methods

For each menstrual group (follicular, luteal, and postmenopausal women), a nested, within person ANOVA was used to test for assay reproducibility over the 4 months. Data were analyzed on the logarithmic scale (base 10) to reduce the dependence of the SD of the response on the mean so that variance can be assumed to be unrelated to subject. Variance components methods were used to model the total variability in the laboratory measurements. Estimates of the variability among women in a given menstrual group (σ^2_a), of assay variability among months for a given woman (σ^2_b), and of assay variability associated with different aliquots on the same month for the same woman (σ^2) were obtained from the SAS procedure Proc NESTED for a nested random effects ANOVA (16); a detailed explanation of this analysis has been provided (7). With y_{ijk} denoting the mean of the log assay measurement over duplicates for women I = 1,2,3,4,5 at month j(I) = 1,2,3,4 on aliquot k(ij)= 1,2, the model is

$$y_{ijk} = \mu + a_i + b_{j(I)} + \epsilon_{k(ij)}$$

where a_i, $b_{j(I)}$, and $\epsilon_{k(ij)}$ are independent variables each with mean zero and respective variances σ^2_a, σ^2_b, and σ^2. From the variance estimates, we computed an estimate of the ICC [as $\sigma^2_a/(\sigma^2_a + \sigma^2_b + \sigma^2)$; i.e., the percentage of the total variability explained by hormone differences among the women] and calculated an estimate of the assay CV. These CV differ from those customarily reported, which are calculated by repeatedly assaying aliquots from the same specimen. Our figures are estimates of the average CV of the five women in each menstrual phase (7). Measurements from all laboratories are pre-

sented in the same units. To ease interpretation, we present geometric means, calculated by raising the means of the $\log_{10}$ (analyte value) to base 10. Spearman correlation coefficients were used to compare the rank order of urine values between laboratories, and Wilcoxon signed-rank tests were used to compare means. Results for each analyte are presented for follicular, luteal, and postmenopausal women separately.

Results

Reproducibility Study. Geometric mean values of each hormone by menstrual group are presented in Table 1. Because our interest was to quantify reproducibility at different concentrations of analyte, values are expressed as analyte/ml; these can be converted to analyte/12 h by considering the total urine collected.

Measurements over the 4 months varied considerably for all of the women, but no systematic time trends were apparent for any hormone (Fig. 1-5). For illustrative purposes, Fig. 1A depicts urinary estrone levels in follicular phase women, with the leftmost symbol being the mean of the duplicates obtained by GC-MS, and the values plotted at months 1–4 being the means of the replicates for the two separate aliquots obtained by RIA methods on those days. Each symbol corresponds to a different woman. These plots graphically display several sources of hormone variability assessed in this study, namely hormone differences among women, measurement variability over time for a given woman, and variability among aliquots on a given day.

Estrone. Results are plotted according to menstrual group and laboratory in Fig. 1, A-C. In premenopausal women, the range of urinary estrone levels was large; for follicular phase women, in particular, levels varied from 7 pg/ml to nearly 80 pg/ml. Over the course of the study, RIA measurements in follicular and luteal women were generally consistent in both laboratories, although measurements were more variable in laboratory 2. Among postmenopausal women, estrone values were low in all instances and were below the limit of detection in laboratory 1 for one woman. (Table 1).
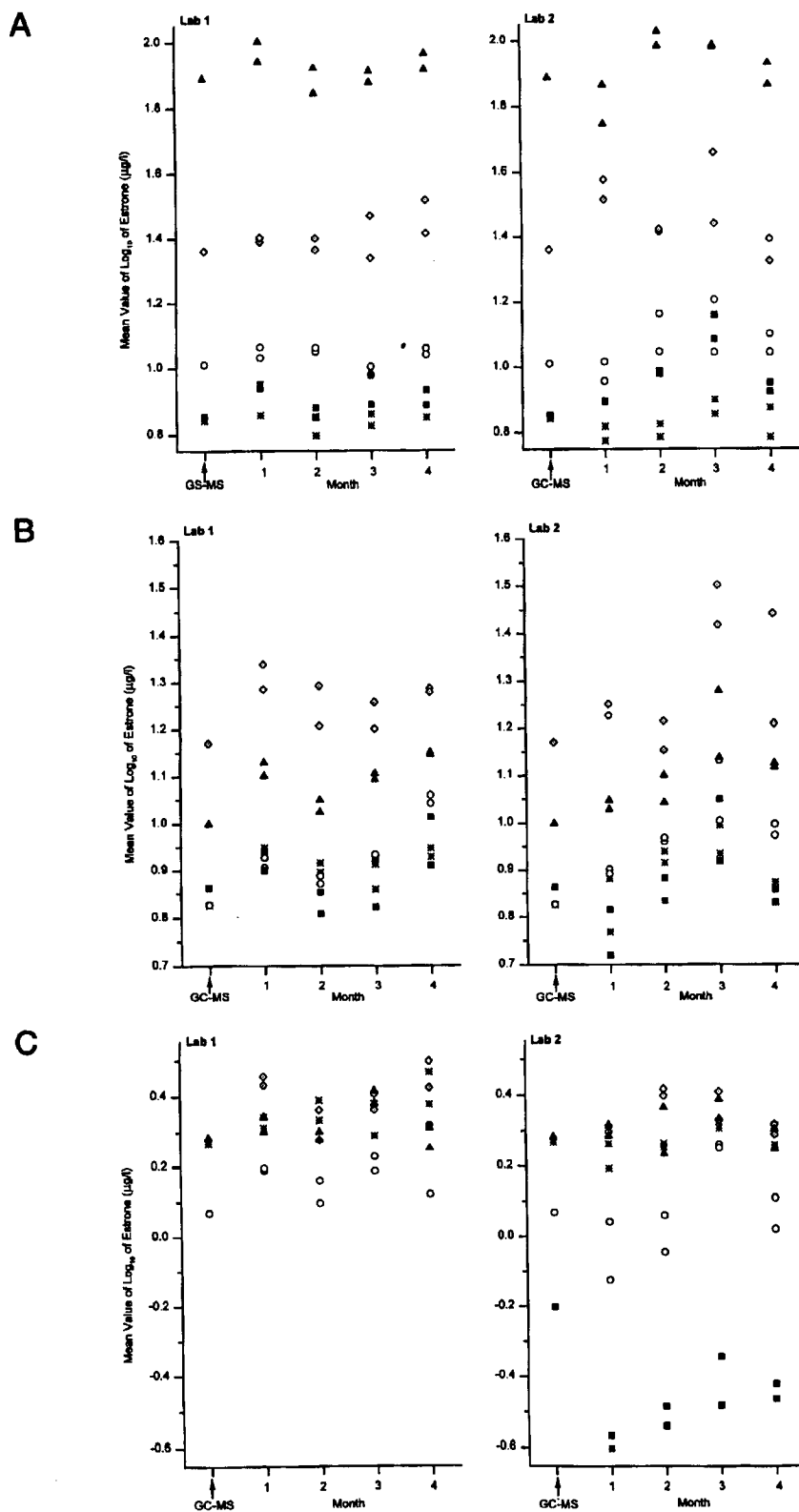
Fig. 1. GC-MS and RIA values of urinary estrone. The leftmost symbol represents the mean of the duplicate values obtained from the GC-MS laboratory. Each of the following symbols represents the mean of the duplicate RIA readings obtained each month. Different symbols distinguish the five follicular phase women (*A*), the five luteal phase women (*B*), and the five postmenopausal women (*C*).

Table 2 Assay CV[a] according to menstrual phase		
	Laboratory	
	1	2
Estrone		
Follicular	10.6	20.8
Luteal	12.7	22.3
Postmenopausal	13	19.3
Estradiol		
Follicular	12.0	34.6
Luteal	11.6	18.9
Postmenopausal	11.7	20.1
Estriol		
Follicular	8.3	33.9
Luteal	12.3	24.6
Postmenopausal	20.0	19.2
Pregnanediol glucuronide		
Follicular	14.7	17.3
Luteal	46.9	8.6
Postmenopausal	8.6	9.5
Estrone glucuronide		
Follicular		15.5
Luteal		10.6
Postmenopausal		22.0

[a] Values shown are $100 \times 2.303 \times (\sigma^2_b + \sim^2_c + \sigma^2/2)^{1/2}$.

Table 3 ICCs[a] by menstrual group and laboratory		
	Laboratory	
	1	2
Estrone		
Follicular	98.7	95.5
Luteal	86.9	75.3
Postmenopausal	85.5	93.3
Estradiol		
Follicular	98.3	89.5
Luteal	93.7	83.1
Postmenopausal	95.1	93.9
Estriol		
Follicular	97.5	85.8
Luteal	96.1	74.8
Postmenopausal	98.6	91.3
Pregnanediol glucuronide		
Follicular	90.3	92.4
Luteal	65.9	84.6
Postmenopausal	76.3	91.8
Estrone glucuronide		
Follicular		96.9
Luteal		92.1
Postmenopausal		90.4

[a] Values were calculated as follows: ICC $= (\sigma^2_a/\sigma^2_a + \sigma^2_b + \sigma^2)$.

In laboratory 1, assay CV were <15%, (Table 2), and hormone differences among women accounted for most of the variability in the measurements with ICCs of 85% or higher (Table 3). In laboratory 2, the RIA results were not as consistent, with CV being ~20% in all menstrual groups. However, for follicular phase and postmenopausal women, differences among women were large, and the resulting ICCs were >90% (Table 3).

Estradiol. RIA measurements did not show any trends over the course of the study in both laboratories, although measurements were more variable in laboratory 2, where differences of 15% or more were not uncommon (Fig. 2, A-C).

In laboratory 1, the assay CV was ~12% in all groups (Table 2), and hormone differences among women accounted for nearly all of the variability in the measurements (Table 3). Assay measurements from laboratory 2 were not as consistent from one month to the next, with CV ranging from 20–35%; yet, the ICCs were generally high (Table 3).

Estriol. Results for estriol varied considerably between laboratories (Table 1). As with estrone, the range of urinary estriol levels in follicular phase women was quite broad (a 10-fold difference in values). No systematic trends were observed in results from either laboratory 1 or 2 over the course of the study (Fig. 3, A-C).

In laboratory 1, estriol results were reproducible for premenopausal women, but varied for postmenopausal women, where the assay CV was 20% (Table 2). For all menstrual groups, hormone differences among women explained most of the variability in the measurements with ICCs >95% (Table 3). Estriol measurements from laboratory 2 were not consistent from month to month in all menstrual groups, with CV ranging from 19–34% (Table 2); nevertheless, ICCs for follicular phase and postmenopausal women were high (Table 3).

Pregnanediol Glucuronide. Pregnanediol glucuronide was not measured by GC-MS. RIA values from laboratory 2 were about twice those from laboratory 1 (Table 1), but no trends in measurements were observed for either laboratory over the course of the study (Fig. 4, A-C).

Unlike the estrogen assays, the RIA for pregnanediol glucuronide in laboratory 1 was problematic. Assay variability was high for luteal phase women (CV >40%; Table 2), and ICCs for the luteal phase and postmenopausal groups were below 80% (Table 3). In laboratory 2, assay results for pregnanediol glucuronide were, for the most part, reproducible, with CV <10% for luteal and postmenopausal women, and the degree of variability among women was high relative to laboratory variability (ICCs were 90% or higher).

Estrone Glucuronide. Urinary estrone glucuronide was measured by RIA in laboratory 2 only; like the other urinary estrogens, levels among follicular phase women spanned a wide range of values. GC-MS results were not available. No systematic trends over the 4 months of the study were observed (Fig. 5, A-C). Assay results were reproducible for premenopausal women (Table 2), and hormone differences among women accounted for most of the measurement variability in all of the menstrual groups where ICCs were 90% or higher (Table 3).

Validity Study. Figs. 6-8 plot the mean of each woman's 16 RIA measurements for estrone, estradiol, and estriol, respectively, against the mean of the corresponding duplicate GC-MS values; Spearman correlation coefficients are presented in Table 4. Hormone means by menstrual group are also provided in Table 1, along with Ps from Wilcoxon signed-rank tests comparing means from the RIA laboratories to the GC-MS laboratory. Two points must be borne in mind when interpreting the plots. First, the RIA value for each woman is the mean of the 16 assays done blindly over the 4 months of this study, making it a particularly precise estimate of that person's hormone level. Second, the GC-MS duplicates are not blinded results.

Estrone. For estrone (Fig. 6), values from laboratories 1 and 2 were well correlated with GC-MS values, with coefficients of 0.70 or higher (Table 4), but RIA levels were significantly higher (Table 1).

Estradiol. Estradiol (Fig. 7) values were highly correlated between laboratories, particularly in premenopausal women where coefficients were 0.80 or higher (Table 4). Although the mean estradiol levels from all of the laboratories were similar,
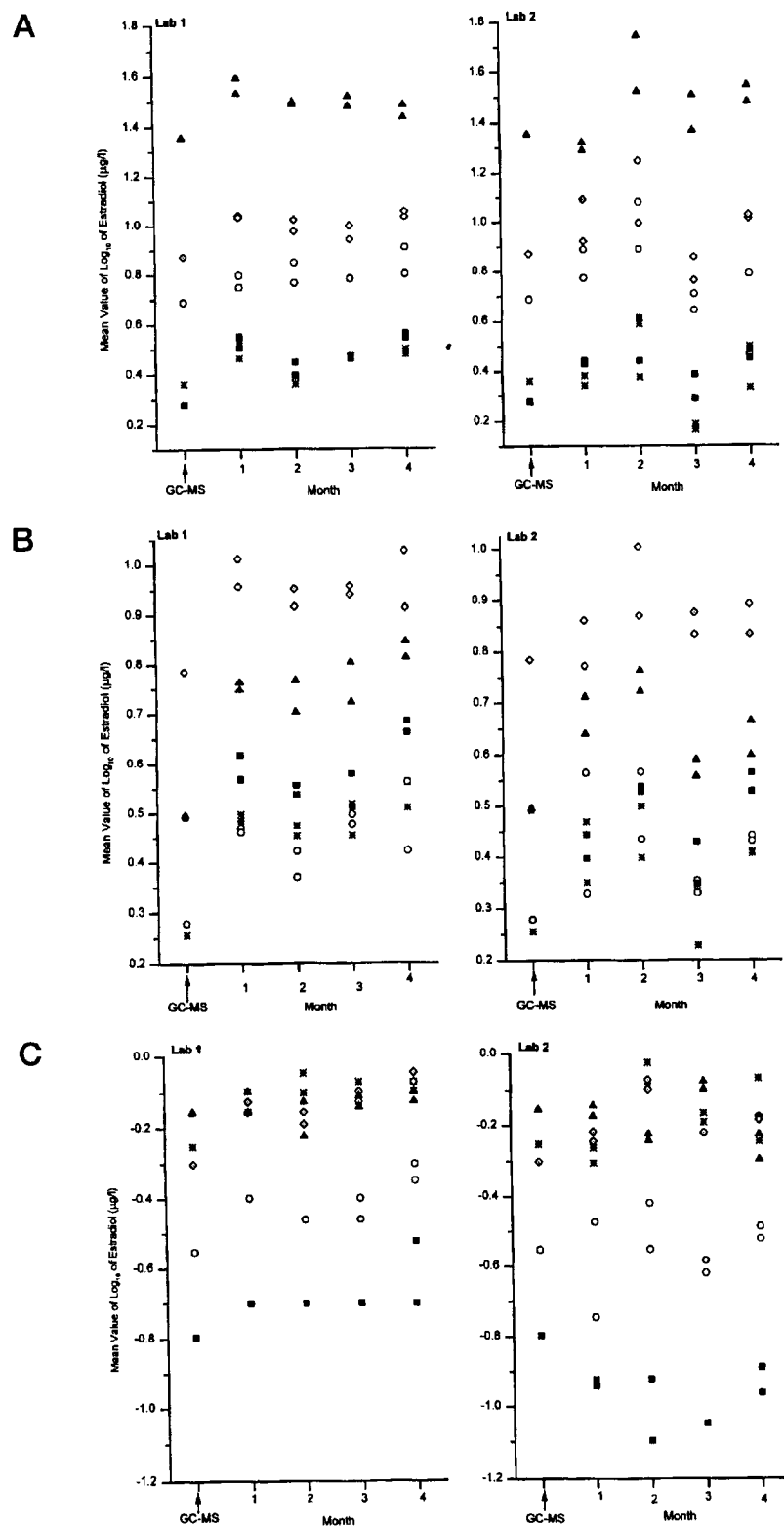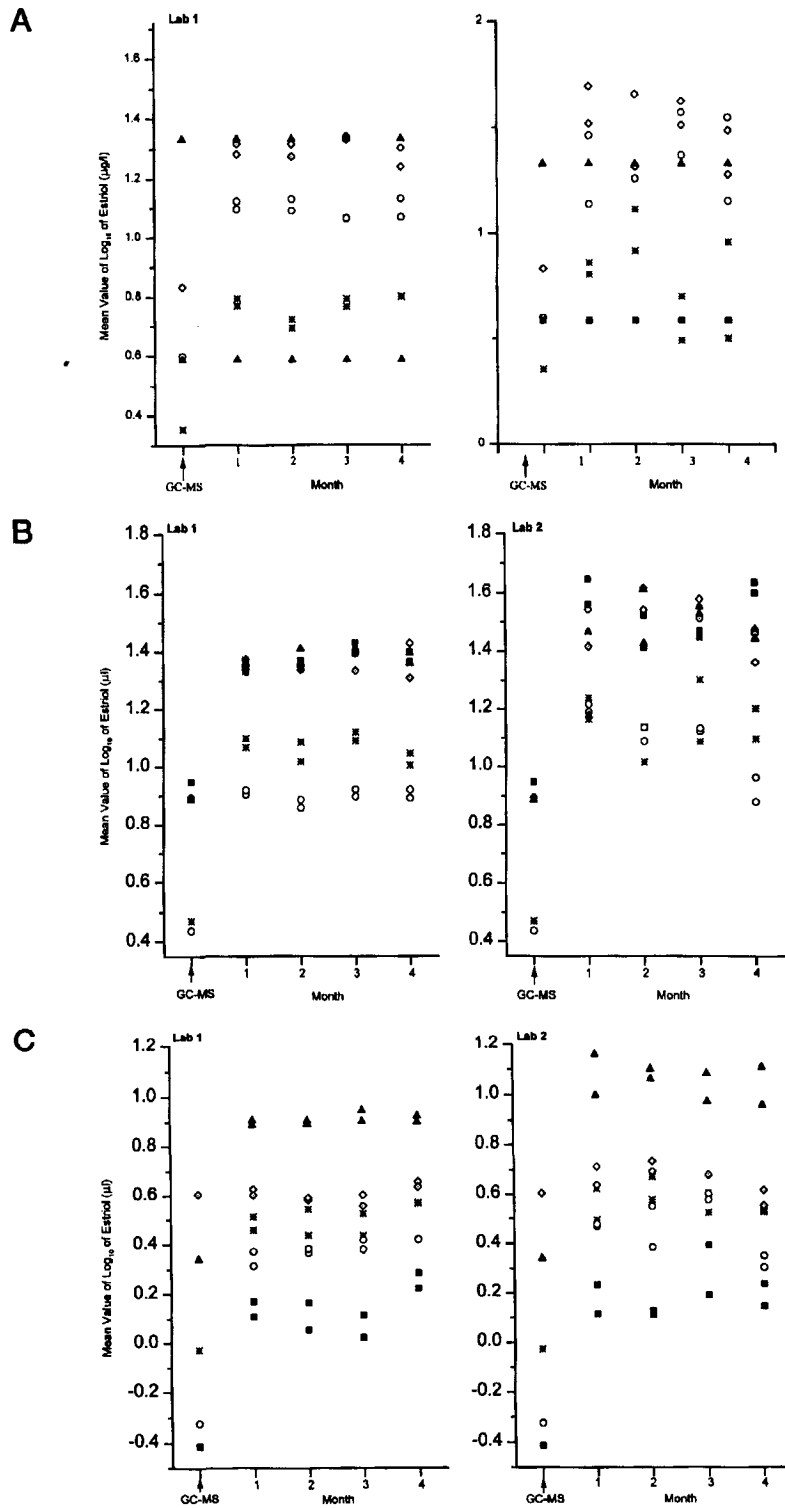
Fig. 2. GC-MS and RIA values of urinary estradiol. The leftmost symbol represents the mean of the duplicate values obtained from the GC-MS laboratory. Each of the following symbols represents the mean of the duplicate RIA readings obtained each month. Different symbols distinguish the five follicular phase women (*A*), the five luteal phase women (*B*), and the five postmenopausal women (*C*).

Fig. 3. GC-MS and RIA values of urinary estriol. The leftmost symbol represents the mean of the duplicate values obtained from the GC-MS laboratory. Each of the following symbols represents the mean of the duplicate RIA readings obtained each month. Different symbols distinguish the five follicular phase women (*A*), the five luteal phase women (*B*), and the five postmenopausal women (*C*).
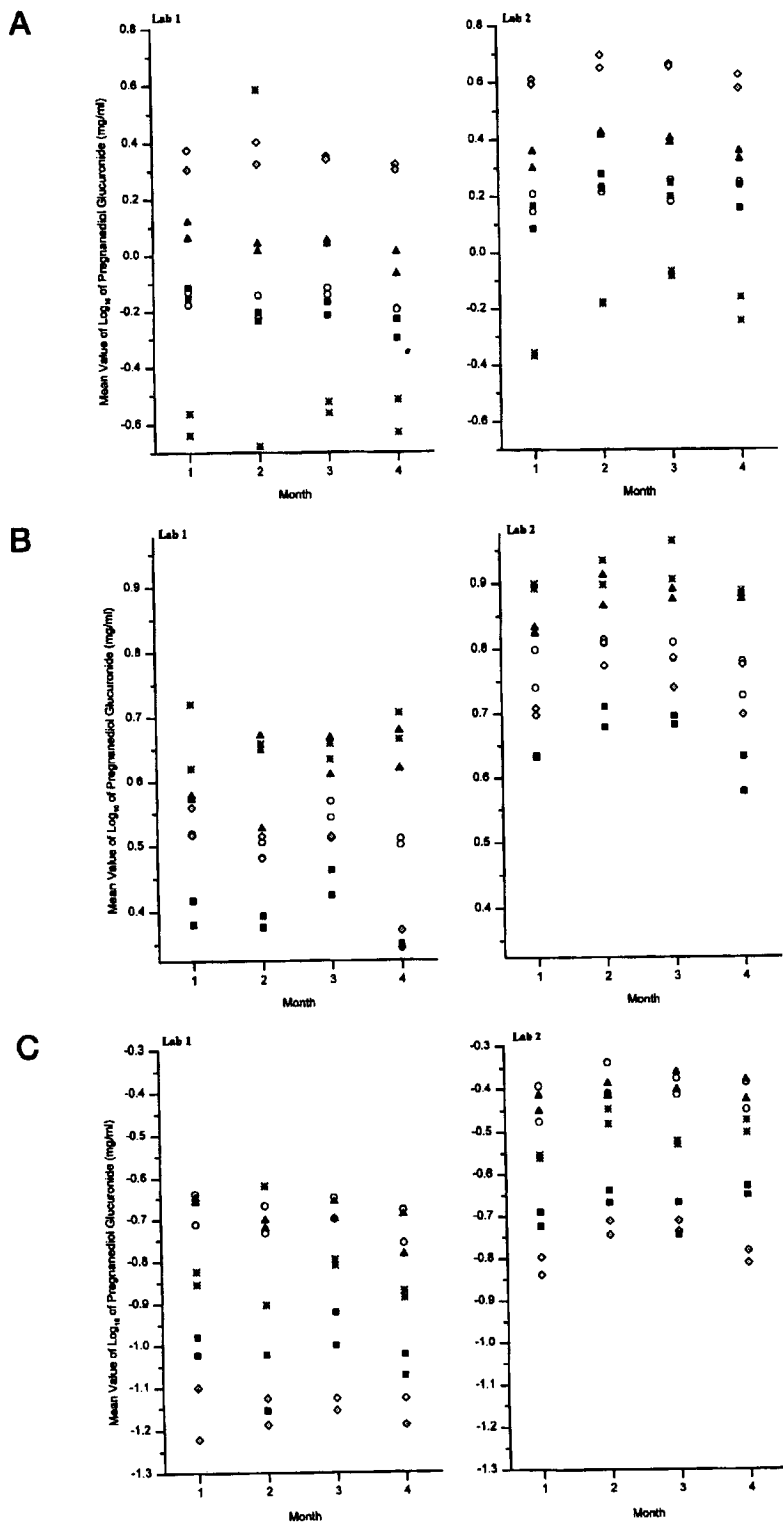
Fig. 4. RIA measurements of urinary pregnanediol glucuronide. Each symbol represents the mean of the duplicate RIA readings obtained each month. Different symbols distinguish the five follicular phase women (*A*), the five luteal phase women (*B*), and the five postmenopausal women (*C*).

Fig. 5. RIA measurements of urinary estrone glucuronide. Each symbol represents the mean of the duplicate RIA readings obtained each month. Different symbols distinguish the five follicular phase women (A), the five luteal phase women (B), and the five postmenopausal women (C).
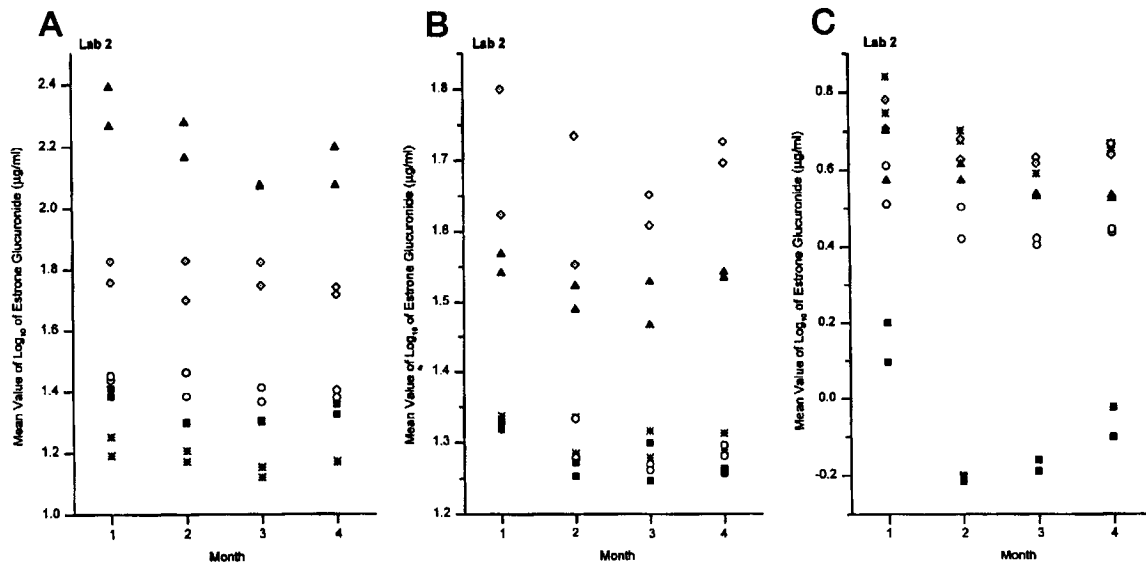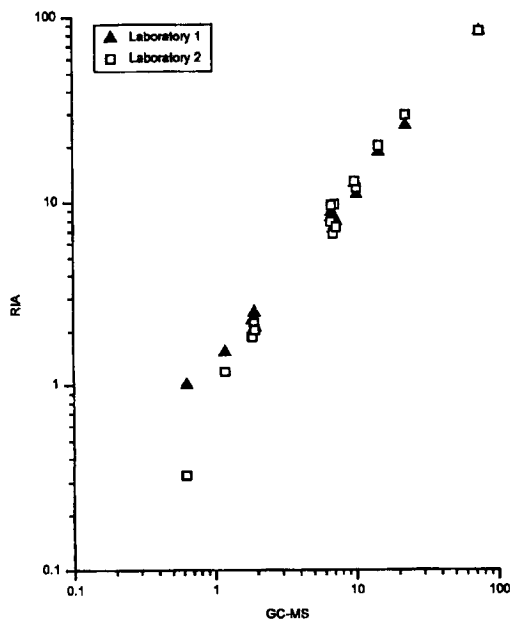

Fig. 6. Plot of RIA and GC-MS measurements of urinary estrone. All measurements are plotted on a $\log_{10}$ scale.

GC-MS measurements were significantly lower than RIA values for all menstrual groups (Table 1).

Estriol. For estriol (Fig. 8), RIA and GC-MS values were highly correlated, but, typically, the RIA measurements were two to three times higher than the GC-MS values (Table 4). RIA results from laboratory 2 tended to be higher than levels reported from laboratory 1 (Table 1).

Discussion

Early breast cancer studies of endogenous hormones measured urinary excretion of estrogens and their metabolites (17), but with advances in assay methodology, measurement of circulating hormones has come to dominate research efforts. Urine collection, however, may be preferable in large population-based surveys where the requirement of a blood draw could sharply reduce study participation rates. Urine has further advantages, including integrating serum levels of hormones that may not have consistent secretion patterns, and, being noninvasive, it is useful for the study of day-to-day hormone fluctuations in a woman's menstrual cycle.

For most women, the catechol estrogens, particularly the 2-hydroxyestrone metabolites, are the most abundant (but relatively unstable) of the urinary estrogens, followed by estrone glucuronide. Although there is no consensus as to which urinary estrogen metabolite best reflects circulating estradiol (2), estrone glucuronide correlates with serum estradiol levels (1), is fairly stable, and, as we have shown, can be consistently measured with RIA. Pregnanediol glucuronide, the major metabolic product of progesterone, has been shown to be highly correlated with serum progesterone levels with a 1–3 day lag (18).

For RIAs of urinary hormones to be useful in population-based research, it must be demonstrated that the assays are reproducible, that the biological variability of the hormone in the population is large relative to assay variability, and that measurements approximate the true values. We found the degree of reproducibility of the RIAs for urinary estrone, estradiol, estriol, and pregnanediol glucuronide differed between the laboratories. No trends in hormone measurements were observed over the course of this study, indicating that storage effects are minimal. Reproducibility was satisfactory in laboratory 1 for all of the analytes, except pregnanediol glucuronide, and for urinary estradiol and estriol, large hormone differences were observed among the women compared to laboratory variability. In laboratory 2, assay results for urinary estrone, estradiol, and estriol were not as consistent, but the

Table 4 Spearman rank correlations[a] among laboratories		
	Laboratory 1	GC-MS laboratory
Estrone		
Overall		
Laboratory 2	0.98	0.95
GC-MS laboratory	0.93	
Follicular		
Laboratory 2	1.00	1.00
GC-MS laboratory	1.00	
Luteal		
Laboratory 2	1.00	0.70
GC-MS laboratory	0.70	
Postmenopausal		
Laboratory 2	0.90	0.90
GC-MS laboratory	0.70	•
Estradiol		
Overall		
Laboratory 2	0.99	0.96
GC-MS laboratory	0.95	
Follicular		
Laboratory 2	1.00	0.90
GC-MS laboratory	0.90	
Luteal		
Laboratory 2	0.90	0.90
GC-MS laboratory	0.80	
Postmenopausal		
Laboratory 2	1.00	0.70
GC-MS laboratory	0.70	
Estriol		
Overall		
Laboratory 2	0.97	0.99
GC-MS laboratory	0.98	
Follicular		
Laboratory 2	0.90	1.00
GC-MS laboratory	0.90	
Luteal		
Laboratory 2	0.70	1.00
GC-MS laboratory	0.70	
Postmenopausal		
Laboratory 2	1.00	1.00
GC-MS laboratory	1.00	
Pregnanediol Glucuronide		
Laboratory 2		
Overall	0.99	
Follicular	0.70	
Luteal	1.00	
Postmenopausal	1.00	

[a] All Spearman correlations with 0.90 are significant at the $P < 0.05$ level.

Fig. 7. Plot of RIA and GC-MS measurements of urinary estradiol. All measurements are plotted on a $\log_{10}$ scale.

Fig. 8. Plot of RIA and GC-MS measurements of urinary estriol. All measurements are plotted on a $\log_{10}$ scale.

ICCs were generally high, so that these assays may yet be useful in large scale studies. For follicular phase women, the wide range of estrone values accounted for nearly all of the variability in the RIA measurements in both laboratories. That the range of estrone values observed in these women adequately represents the population values (or for that matter, the range of values of any analyte studied) may be questionable, because only five women in each menstrual phase participated in this study. With so small a sample, the precision of the variance component for between person differences is poor.

For estrone and estradiol, RIA measurements in both laboratories were well correlated with each other, as well as with the GC-MS measurements. It is noteworthy, however, that differences in RIA measurements of 15% or more were not uncommon because it has been speculated that as little as a 10% difference in hormone concentrations might be associated with a substantial breast cancer risk. For estriol, absolute values
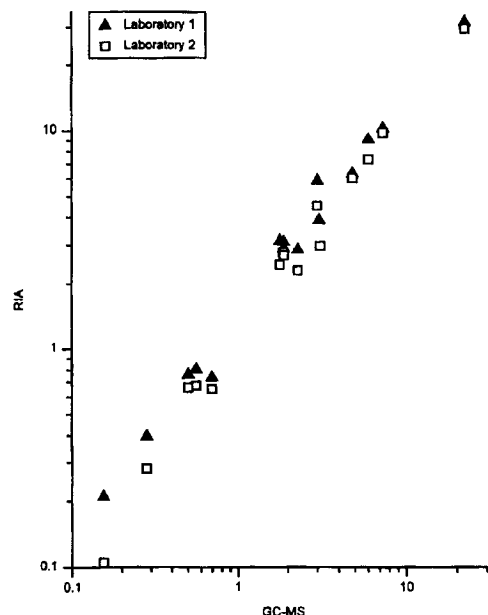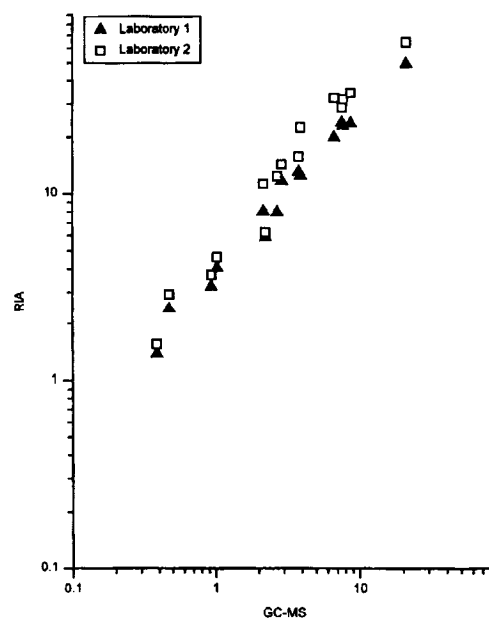
differed considerably among all of the laboratories, with RIA measurements from laboratory 1 being consistently lower than those from laboratory 2, and GC-MS estimates being lower still. Because measurements were proportionately different from one laboratory to the next, the correlations between laboratories for this analyte were high. Unidentified cross-reactiv-

ity in the laboratory assay, unmeasured loss of analyte during hydrolysis before GC-MS, or calibration problems may account for some of these measurement differences.

Epidemiological studies can be designed to overcome imprecision in the laboratory assay by increasing the study size and by batching matched cases and controls for the assays. For instances in which batching is not possible, statistical adjustment should be attempted. The contribution of biological fluctuations within a woman was not addressed in this methodological study because only one specimen was collected from each woman. A study of intrawoman variability would entail sample collections on several days from the same woman. In such a study, the time of day as well as day(s) of the menstrual cycle that urine is to be collected must be specified, because hormones are known to fluctuate widely during the menstrual cycle, to be influenced by circadian and/or diurnal secretion patterns, or to be secreted in a pulsatile fashion.

In summary, we found that the reproducibility of most of the assays for urinary hormones and metabolites differed between the laboratories, yet assay variability was low in comparison to the magnitude of biological variability among women. For postmenopausal women in particular, assays of urinary estrogens may be alternatives to serum assays where hormone levels are close to the limits of assay sensitivity. Current laboratory RIAs for urinary hormones and their metabolites make urine collection a useful and easily obtained resource in large-scale studies attempting to link hormones to cancer risk.

References

1. Wilcox, A. J., Baird, D. D., Weinberg, C. R., Armstrong, E. G., Musey, P. I., Wehmann, R. E., and Canfield, R. E. The use of biochemical assays in epidemiologic studies of reproduction. Environ. Health Perspect., 75: 29–35, 1987.

2. Lasley, B. L., and Shideler, S. E. Methods for evaluating reproductive health of women. Occup. Med., 9: 423–433, 1994.

3. Potischman, N., Falk, R. T., Laiming, V. A., Siiteri, P. K., and Hoover, R. N. Reproducibility of laboratory assays for steroid hormones and sex hormone-binding globulin. Cancer Res., 54: 5363–5367, 1994.

4. Hankinson, S. E., Manson, J. E., London, S. J., Willett, W. C., and Speizer, F. E. Laboratory reproducibility of endogenous hormone levels in postmenopausal women. Cancer Epidemiol. Biomark. Prev., 3: 51–56, 1994.

5. Toniolo, P., Koenig, K. L., Pasternack, B. S., Banerjee, S., Rosenberg, C., Shore, R. E., Strax, P., and Levitz, M. Reliability of measurements of total, protein-bound, and unbound estradiol in serum. Cancer Epidemiol. Biomark. Prev., 3: 47–50, 1994.

6. Cauley, J. A., Gutai, J. P., Kuller, L. H., and Powell, J. G. Reliability and interrelations among sex hormones in postmenopausal women. Am. J. Epidemiol., 133: 50–57, 1991.

7. Gail, M. H., Fears, T. R., Hoover, R. N., Chandler, D. W., Donaldson, J. L., Hyer, M. B., Pee, D., Ricker, W. V., Siiteri, P. K., Stanczyk, F. Z., Vaught, J. B., and Ziegler. R. G. Reproducibility studies and interlaboratory concordance for assays of serum hormone levels: estrone, estradiol, estrone sulfate, and progesterone. Cancer Epidemiol. Biomark. Prev., 5: 835–844, 1996.

8. Ziegler, R. G., Rossi, S. C., Fears, T. R., Bradlow, H. L., Adlercreutz, Sepkovic, D., Kiuru, P., Wahala, K., Vaught, J. B., Donaldson, J. L., Falk, R. T., Fillmore, C., Siiteri, P. K., Hoover, R. N., and Gail, M. H. Quantifying estrogen metabolism: an evaluation of the reproducibility and validity of enzyme immunoassays for 2-hydroxyestrone and 16α-hydroxyestrone in urine. Environ. Health Perspect., 105 (Suppl. 3): 607–614, 1997.

9. Stanczyk, F. Z., Miyakaw, I., and Goebelsmann, U. Direct radioimmunoassay of urinary estrogen and pregnanediol glucuronides during the menstrual cycle. Am. J. Obstet. Gynecol., 137: 443–450, 1980.

10. Magini, A., Pinzani, P., Bolelli, G. F., Bassi, F., Salerno, R., Messeri, G., and Pazzagli, M. Measurement of estrone-3-glucuronide and pregnanediol-3 α glucuronide in early morning urine samples to monitor ovarian functions. J. Biolum. Chemilum., 4: 567–574, 1989.

11. Stanczyk, F. Z., Shoupe, D., Nunez, V., Macias-Gonzales, P., Vijod, M. A., and Logo, R. A. A randomized comparison of nonoral estradiol delivery in postmenopausal women. Am. J. Obstet. Gynecol., 159: 1540, 1988.

12. Cassidenti, D. L., Vijod, A. G., Vijod, M. A., Stanczyk, F. Z., and Lobo, R. A. Short-term effects of smoking on the pharmacokinetic profiles of micronized estradiol in postmenopausal women. Am. J. Obstet. Gynecol., 153: 1953–1960, 1990.

13. Katagiri, H., Stanczyk, F. Z., and Goebelsmann, U. Estradiol in pregnancy. III. Development, comparison and use of specific antisera for rapid radioimmunoassay of unconjugated estriol in pregnancy plasma. Steroids, 14: 224, 1974.

14. Fotsis, T., and Adlercreutz, H. The multicomponent analysis of estrogens in urine by ion exchange chromatography and GC-MS. I: quantitation of estrogens after initial hydrolysis of conjugates. J. Steroid Biochem., 28: 203–213, 1987.

15. Bannwart, C., Adlercreutz, H., Wahala, K., Brunow, G., and Hase, T. Deuterium labeled ethoximes as stable isotope internal standards in the GC-MS-SIM determination of oxo-estrogens in human urine extracts: preliminary results. In: S. Gorog and E. Heftmann (eds.), Proceedings of the Symposium on the Analysis of Steroids, Sopron, Hungary (1987), Advances in Steroid Analysis, 1987, pp. 283–286. Budapest, Hungary: Akademiai Kiado, 1988.

16. SAS Institute, Inc. SAS/STAT User's Guide, Version 6, Ed. 4, Vol. 2. Cary, NC: SAS Institute Inc., 1989.

17. Key, T. J. A., and Pike, M. C. The role of oestrogens and progestagens in the epidemiology and prevention of breast cancer. Eur. J. Cancer Clin. Oncol., 24: 29–43, 1988.

18. Munro, C. J., Stabenfeldt, G. H., Cragun, J. R., Addiego, L. A., Overstreet, J. W., and Lasley, B. L. Relationship of serum estradiol and progesterone concentrations to the excretion profiles of their major urinary metabolites as measured by enzyme immunoassay and radioimmunoassay. Clin. Chem., 37: 838–844, 1991.